11

القياس النفسي للتكييف:
تقييم تعادل القياس عبر اللغات والثقافات

فيتز دراسكو وتاهيرا برويسن
جامعة الينوي

إننا نجد أن معاهدة التجارة الحرة في أمريكا الشمالية (NAFTA) وتشابك الوحدة الأوربية هما مجرد مؤشرين على الاتجاه المتزايد نحو العولمة في الأسواق الدولية. وكمؤشر على وتيرة التغير نجد أن الصادرات التجارية الدولية ارتفعت من (7.382) بليون دولار في سنة 1985 إلى /1.1/ تريليون دولار عام 1996 (باتش 1998). وبالإضافة إلى تشكيل دول لاتحادات اقتصادية كبرى، فقد أصبح شيئاً اعتيادياً لشركات كبيرة في دول مختلفة مثل شركة "كرايسلر" و"ديملربنز" أن تندمج لتجعل نشاطاتها التجارية متكاملة. وكنتجية لعولمة أسواق العالم تصبح الحاجة لفهم الفروق الحضارية بين الناس المتباينين في المشارب والأصول الحضارية أكثر وضوحاً.

وقد كان من المعالم المهمة في التمازج بين الثقافات نشر كتاب "عواقب الثقافة" في عام 1980 لمؤلفه "جرته هوفستيد" الذي دوّن أبعاداً متعددة ومهمة تختلف بموجبها الثقافات. فبالتحليل العاملي لما يزيد عن 116 ألف استبانه أخذت من موظفي شركة "IBM" في "40" بلداً مختلفاً، وجد هوفستيد أن الثقافات تختلف في كلٍّ من مسافات النفوذ Power distance، وتجنب اللامحدودية Uncertainty، والرجولة Masculinity، والفردية Individualism. فمسافات النفوذ تشير إلى

القابلية التي ترى بها المسافات الواسعة بين المستويات الأعلى والأدنى في الهرم الاجتماعي. أما تجنب عدم الوضوح في المواقف فيعكس تجنب المواقف التي تكون النتيجة فيها غير واضحة. والرجولة تعني نظرة أعضاء المجتمع الحضاري إلى أنشطة قيم تعتبر عرفياً أكثر رجولة في ثقافتهم، ولكن البعد المدروس بتمحيص أكثر من غيره بكثير فهو البعد الذي يتناول الفردية والجمعية (الجماعية).

وقد سجل تريانداس (1990-1994-1995) بعض الاختلافات القوية، الواضحة التي ينظر فيها الناس، في بعد الثقافات الفردية والجماعية في العالم؛ في بعض الثقافات الفردية مثل الولايات المتحدة الأمريكية وكندا وكثير من دول غرب أوربا، يركز الأفراد على الذات. فالأفراد يؤكدون على الحاجات والرغبات لدى الفرد قبل حاجات الجماعة ورغباتها. ويفترض بالأفراد أن يكونوا مستقلين بأنفسهم، ويتقرر السلوك على الأغلب بالأهداف الفردية. وكنتيجة نجد أن الشعور الجماعي ليس له تأثير كبير على الأفراد. ومن الناحية الأخرى نجد أمريكا الجنوبية وأمريكا الوسطى وآسيا وأجزاء عديدة في العالم المتطور جماعية النزعة. ومن وجهة نظر تريانداس (1990) يؤكد أولو النزعة الجماعية على وجهات النظر والحاجات والأهداف المشتركة أكثر من حاجات الذات وهنا تنحو رغبات الفرد لأن تكون تابعة للحاجات المشتركة. وتقرر نمط السلوك بشكل كبير بالعرف الاجتماعي والواجبات أكثر من الرغبات الفردية، كما أن التناسق الاجتماعي المشترك يعد ذا قيمة عالية.

وعلى الرغم من أن الباحثين في علم النفس التنظيمي في الصناعة، وباحثي إدارة الموارد البشرية يدركون أن الفروقات الثقافية لا يمكن تجاهلها في المنظمات متعددة الجنسيات، فإن البحث الإجرائي في هذه الموضوعات قد تقدم بشكل بطيء بسبب المشكلات النظرية والعملية العديدة. فالحصول على فرصة لإجراء البحث في المنظمات متعددة الجنسيات باستخدام التسهيلات الدولية أكثر صعوبة من الحصول عليها في مجالات الأعمال الأمريكية. وبعد الحصول على الإذن بإجراء

البحث يتبين أن التكلفة تكون عالية بشكل لا يشجع على إجراء البحث. إلا أنه قد تكون أشق عقبة يتوجب تخطيها في محاولة إجراء البحث الذي يتناول موضوع دراسة الثقافات سوية بشكل شامل باستخدام الأساليب السليمة، تكمن عند مرحلة تكييف الأدوات المستخدمة، فقد قيل عن التكييف إنه: "قد يكون أكثر أنواع الأحداث تعقيداً بين الأحداث التي ظهرت حتى الآن منذ تطور الخليقة". (ريتشاردز 1953 ص250)؛ وعلى الرغم من أن عبارة ريتشاردز لا شك قد تكون مبالغاً فيها بالطبع، إلا أن هذا الفصل، بل هذا الكتاب بأكمله هو عبارة عن تقرير حول التحديات الكامنة في إجراءات التكيف والترجمة.

هناك العديد من الخطوات التي يحتاج الباحثون لاتخاذها لضمان التعادل في النوعية والقياس بين الأدوات التي يتبنونها لإجراء بحثهم، وهذه الخطوات شديدة الأهمية، إذ بدونها قد تهدر كميات هائلة من وقت البحث ومن الجهد المبذول في سبيله ومن الأموال دون طائل. يبدأ مسار تكييف الأداة المستخدمة بإجراء استبيان شامل يوضع أساساً في العادة بلغة معينة من اللغات المصدر (اللغة المترجم منها)، وعند إجراء البحث في المقارنة بين الثقافات كثيراً ما يكون ضرورياً أن تجرى تكييفات للأداة في لغة أو أكثر من اللغات المتلقية للترجمة، أو اللغات التي يجري التركيز عليها. والشكل الأفضل يكون باستخدام مجموعتين من المترجمين لهذا الغرض، يقوم أفراد المجموعة الأولى من المترجمين بشكل إفرادي بترجمة الاستبيان من لغة المصدر إلى اللغة المتلقية وبعد ذلك يقوم أفراد هذه المجموعة بالاجتماع وتسوية الاختلافات في ترجماتهم.

وبعد أن يتم الاتفاق على الشكل النهائي يعطى الإستبيان لكل مترجم من الفئة الثانية فيقوم هؤلاء المترجمون بدورهم، كل بمفرده، بإعادة ترجمته من اللغة المتلقية إلى اللغة الأصل، وبعد أن يسوي هؤلاء المترجمون بدورهم اختلافاتهم يقارن

الاستبيان الأصلي مع الورقة المعاد ترجمتها وفي نهاية هذه الإجراءات الطويلة المعقدة يكون من دواعي الارتياح والطمأنينة أن تظهر النسخة الأخيرة للترجمة متشابهة بشكل عالٍ مع النسخة الأصلية من الاستبيان.

إلا أنه كثيراً ما يكون هناك تفاوت في أحوال عديدة بين النسخة الأصلية والنسخة المعاد ترجمتها؛ فمثلاً من خلال العديد من المحاولات التي تبذل للتكييف واجهتنا مراراً صعوبات في ترجمة التعابير الأمريكية المصطلح عليها اجتماعياً إذ عندما نصف عملا على أنه «مريح» يفهم الأمريكيون ذلك إلا أنه في لغة "المراثي Marathi" ترجمت هذه العبارة بكلمة "يؤدي بشكل سهل" وكذلك يمكن لنا أن نرى أن وصف عبارة (waste of time) أي «هدر الوقت» في اللغة الإنكليزية قد يترجم على أنه (وقت زملاء العمل المهدور).

وبالإضافة إلى هذه التعابير المتعارف عليها يمكن أن تقابل الصعوبات أيضاً عند ترجمة قطعة بسيطة نسبياً من النثر، واللغة الإنكليزية لغة فيها الكثير من المترادفات. وأحياناً تصاغ المقاييس بوضع السؤال ذاته باستخدام أشكال متعددة من الكلمات أو العبارات المتشابهة، وحيث إن المترادفات كثيراً ما تنشأ لغوياً "عندما يكون المفهوم ذا أهمية في ثقافة معينة (فمثلاً لدى الإسكيمو أكثر من مئة كلمة لوصف الثلج، فإن الثقافات الأخرى قد لا يكون لديها إلا وصف واحد للكلمة). وكمثال في بحثنا الذي يتناول التوتر بشكل عام، يسأل مقياس (SIG) الذي وضعه سميث وساديمان ومكراري في عام (1962) المستجيبين أن يذكروا إذا كان عملهم "مزحوماً" أو "داعياً للإزعاج"، إلا أنه في لغة "المراثي" لا تجد سوى كلمة "مزعج" كأقرب وصف يمكن ترجمة المفهوم إليه. وعلى ذلك، عند استخدامنا لمقياس (SIG) بشكله الكامل كان لا بد لنا أن نكتفي باستخدام كلمتي "مزعج" و"مزعج جداً" عند ترجمة هاتين الكلمتين. وعند حدوث تباينات وتغيرات في المدلول في الترجمة بين اللغة مصدر النص واللغة المتلقية عندئذٍ يجب علينا في هذه الحالة أن نكرر إجراء

الترجمة إلى أن يتم الحصول على الامتزاج السليم. وطبعاً "لا تزال هناك معضلة أخرى في سر التكييف وهي تقرير ما يكون عليه الامتزاج السليم (الوافي)". وعلى الرغم من وجود درجة عالية من التشابه بين النص الأصلي والترجمات المعادة ترجمتها من الاستبيان إلا أن ذلك لا يكفل التعادل بين النص في اللغة الأصلية والنص في اللغة المترجم إليها. فمثلاً يمكن للترجمة المشوشة أن تعاد إلى ما يشبه النص الأصلي إذا تمكن الذين يعيدون الترجمة من التكهن بما يعنيه المترجمون الأولون. وهناك مشكلة أقوى سببها أن الأفراد الذين يتقنون لغتين لا يشبهون في أي من اللغتين أولئك الذين يتقون لغة واحدة فقط في أي من الجانبين (لاندر وايرفنغ وهوروميز 1960). وأهمية هذه النقطة أكثر من مجرد أكاديمية، فعندما درسنا نوعية ترجمة لمقياس يقيم الرضى بالعمل باستخدام ذوي اللغتين (هولين ودراتكو وكريوكار 1982) لم نجد الكثير من دلائل عدم التكافؤ بين اللغتين الإنكليزية والإسبانية (مجرد 3 بنود من أصل 72 بند كانت التي ظهر فيها الاختلاف في القياس بين اللغات) ولكن فيما بعد عندما قارنا الشكلين ذاتهما في مقياس "الرضى بالعمل" بتطبيقه على متكلمي اللغة الإنكليزية وحدها ومتكلمي اللغة الإسبانية وحدها (دراسكو وهولين 1988) وجدنا كثيراً من الاختلافات عبر اللغتين (حوالي ثلث بند المقياس). وبالرغم من متابعة مسار التكيف المذكور أعلاه بدقة هناك على الأقل ثلاثة أسباب لكون تحليل خواص القياسات لبند المقياس قد تكشف عن اختلافات عبر اللغات. فالسبب الأول والأكثر وضوحاً هو أن البنود لم تكن مترجمة بشكل سليم. فمثلاً الفرق بين كلمة حقل وكلمة مزرعة تربية أبقار أمريكية بيّن، واختيار كلمة في اللغة التي تجري الترجمة إليها تقارب كلمة مزرعة الأبقار تماماً شيء صعب جداً. ثانياً قد تكون بعض المفاهيم المألوفة في إحدى الثقافتين صعبة أو مستحيلة الفهم على أعضاء ثقافة أخرى. فمثلاً عبارة "ما حك جلدك مثل ظفرك" (Do your own thing) وعبارة: "لا يعيش الإنسان إلا مرة واحدة" (You only live once) تضعان التركيز على الذات التي تفهم، بل وينادى بها أيضاً، في البلدان ذات

الثقافات فردية الاتجاه، إلا أن مثل هذا التجاهل للأصدقاء والعائلة يمكن أن لا يفهم بسهولة من قبل أفراد ثقافات تركز على المجموعة أكثر من تركيزها على الفرد. وأخيراً يعتقد أحياناً أن الأفراد القادمين من ثقافات مختلفة يميلون إلى استخدام مقاييس الإجابة بشكل مختلف. وبشكل محدد ينحو بعض الأفراد إلى تجنب استخدام الشكل الأقصى لمقياس الإجابة (مثلاً الاختيار بين /1/ و/7/ من مقياس مثل مقياس الإجابة "ليكرت" المكون من 7 بنود)، بينما يميل أفراد ثقافات أخرى إلى إعطاء تقييمات مبالغ فيها (هوي وتزياندليز 1989، تريانداريز 1972) فمثلاً وجد ماران وكامب وماران (1992) أن متكلمي اللغات الإسبانية أكثر ميلاً لاستخدام إجابات متطرفة وأن يتفقوا مع التقارير أكثر من الشعوب غير الإسبانية فهم يعكسون جزئياً فروقاً في الأعراف الاجتماعية المتعلقة بالموافقة.

تجعل هذه المشكلة المرء يرى بشكل جلي أن مقارنة إجابات الأشخاص الذين يتكلمون لغات مختلفة وينشؤون في ثقافات مختلفة أمر عسير جداً - إذ كيف يمكن للباحثين أن يقروا إجرائياً ما إذا كانت البنود والمقاييس المعطاة لمثل هذه المجموعات المتباينة من الناس تقيس بشكل متعادل فعلياً، هدف هذه الفصل هو توضيح منطلق واحد لمثل هذه التحليلات فنحن نستخدم نظرية الإجابة وفق البنود (IRT) لدراسة ما إذا كانت العلاقة بين احتمال اعتماد بند والخاصية الخفية الكامنة التي يقيسها المقياس هي متماثلة لدى المجموعات كافة.

نظرية إجابة البنود Item Response Theory

نقدم في هذا الفصل وصفاً موجزاً لهذه النظرية (IRT) يقدم هاملتون وسواميناثان (1985) وبيكر (1992) وهولين ودرازكو (1983) تفصيلات أكثر توسعاً حولها.

لنفرض أن U1 وU2 وUn... تصف عدداً (ن) (n) من البنود على مقياس ما. لتبسيط الأمور لا نعتبر إلا نماذج الإجابات ذات الدرجات المأخوذة ثنائياً، هنا (Ui=1) من أجل إجابة إيجابية لبند موضوع بشكل إيجابي (مثلاً إجابة بنعم لبند

يسأل: "هل يبعث عملك على الرضى في نفسك؟") أو جواب بالنفي لبند مصوغ بشكل سلبي مثلاً: جواب بالنفي على سؤال يقول: "هل عملك جميل؟" المتغير الثنائي Ui=0 لجواب بالنفي مصاغ إيجابياً (مثلاً جواب كلا لسؤال يسأل: (هل عملك رائع؟") أو جواب إيجابي لسؤال سلبي الصيغة: ("هل عملك يقودك للإحباط؟") وقد أصبحت النماذج المتعددة الأبعاد التي تسمح بتصنيف الإجابات في زمر منظمة (مثلاً نموذج سمجيما 1969) متدرج الإجابات، أو الزمر الاسمية (مثلاً النموذج الاسمي الذي وضعه "بوك" (1972) شائعة ومرغوبة على نطاق واسع (عدد آذار 1995 من مجلة القياسات النفسية التطبيقية الذي كان مكرساً للنماذج متعددة الأبعاد) ولكن ذلك خارج نطاق بحثنا الآن.

وكذلك لا ننظر إلا إلى نماذج الإجابة للبند التي تكون أحادية الأبعاد (ولكن يمكن ملاحظة أن عدد كانون الأول 1996 من مجلة القياس النفسي التطبيقي كان مكرساً لنماذج متعددة الأبعاد من الإجابة للبنود). بالنسبة للنماذج أحادية الأبعاد تستخدم قيمة - (ثيتّا) المقياسية في كثير من الأحيان لتعني الخاصية الخفية الكامنة المقيّمة موجب (n)"ن" على المقياس، ووظيفة إجابة البند /1RF/ التي تسمى أحياناً المنحنى المميز للبند هي أساسية في هذه النظرية، فهي تعطي احتمال إجابة إيجابية (أي أن Ui=1) للبند كوظيفة "θ" ويرمز إليها اعتيادياً بـ "P (θ)". في هذا الفصل نركز على النموذج التوريدي ثنائي الأبعاد الذي يستخدم الشكل الرياضي.

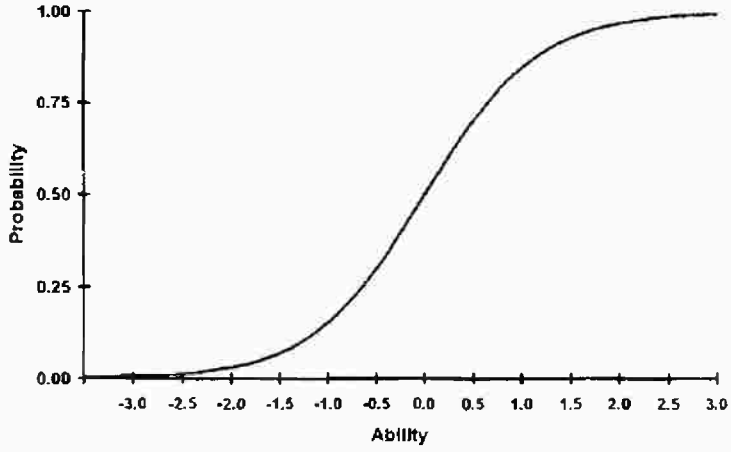
$$P_i(\theta) = \frac{1}{1 + \exp[-Da_i(\theta - b_i)}$$

في المعادلة السابقة تكون D مجموعة ثابتة تعادل 1.702 لأسباب تاريخية (أي بحيث أن يطابق النموذج التوريدي بدقة وظيفة إجابة البندlogistic الطبيعية) =Ex (ExP) [X] حيث الثابت الرياضي يساوي تقريباً 2.718 في هذه المعادلة تكون ai وbi نواظم البند التي تكون ذات أهمية جوهرية بالنسبة لنا. ومؤشر التباين (التمييز) للبند ويرمز إليه بـ (ai) يدل على حدة ميل إجابة البند، وإجابة الميل التي

ترتفع بزاوية حادة الميل في فترة ما تمكننا من التمييز بدقة بين المجيبين الذين تكون لديهم "θs" أدنى أو المجيبين الذين تكون لديهم "θs" أعلى بعض الشيء. ومؤشر صعوبة البند bi مؤشر مكاني حيث θ=bi يجب الأخذ بعين الاعتبار أن Pi(θ) = 0.5 بحيث تكون bi النقطة على الاستمرارية الخفية الكامنة حيث يكون للمجيبين فرصة 50٪ للإجابة.

يعرض الشكل 1.11 وظيفة الإجابة للبند بالنسبة لبند افتراضي في مستويات q الدنيا تدل وظيفة إجابة البند على أن المجيبين سيكون لديهم احتمال تبني البند يقرب من الصفر، ويكون للمجيبين احتمالات إجابة إيجابية تقترب الواحد لدى قيم q العالية وبالتالي يقدم البند الافتراضي تمييزاً جيداً بين المجيبين الذين تكون لديهم قيمة q منخفضة بشكل معتدل ومرتفعة بشكل معتدل.

يسلك مسار استخدام نظرية إجابة البند لتقييم تعادل القياس لبند ما بين مجموعتين ثقافتين الاتجاه التالي: أولاً يجب جمع عينات ممثلة لذلك من كل من أفراد كل من الثقافتين، ثم نقدر نواظم البند بشكل مستقل للمجموعتين.



الشكل رقم 11-1

منحنى مميز لبند مفترض

وبعد وصل قياسات θ للمجموعتين (أي وضع تقديرات للنواظم على المقياس ذاته) يمكننا أن نقارن ما إذا كانت إجابة البند للثقافة الأولى تختلف بشكل ذي دلالة عن وظيفة إجابة البند للثقافة الثانية. يتم هذا التقييم باختبار أهمية متعدد التباينات يقارن تقديرات ai وb من عينة المستجيبين من الثقافة الأولى مع ما تم تقديره لـ ai وb من الثقافة الثانية. في الفصل التالي تم وصف كل خطوة من التحليل الذي استخدمناه لدراسة التعادل بين الثقافات في المقاييس المكيفة بتفصيل أكبر.

خطوات في تحليل المعلومات:

البُعدية: في الجزء الأكبر استخدمت الدراسات التي تبحث في تعادل القياسات عبر الثقافات عن طريق نظرية إجابة البند، نماذج وحيدة (أحادية) الأبعاد (أي نماذج تظهر فيها - ذات تدريجات وليس "vector" في البحث الجاري وصفة في هذا الفصل نستخدم نموذج "logistic" ذات ناظمين يكون أحادي الأبعاد وبذلك من الضروري اختبار الدرجة التي تنطبق فيها المعلومات على الافتراض أن هناك طرقاً عديدة لدراسة البعدية "Dimensionality". والظاهر أن هناك طرقاً عديدة لتقييم البُعدية بعدد الأشخاص الذين يقومون بعمل القياس النفسي. قدم "هاتي 1985" سرداً مفصلاً للطرق التي يمكن باستخدامها قياس البُعدية.

والطريقة التي نستخدمها عادة لقياس البعدية طريقة بسيطة؛ تحسب التناسقات ثلاثية الوظائف tetrachoric من أجل البند المدرج بشكل ثنائي وبعد ذلك تحلل المصفوفة(Matrix) الارتباط بالتحليل العاملي للمحور الرئيس، وبشرط أنه لا يوجد من العوامل ما هو متطرف أكثر مما يجب (أي أن كافة المستجيبين تقريباً يجيبون سلبياً، أو كل المستجيبين يجيبون إيجابياً) وتقدم أول قيمة "eigen" التي تكون عالية نسبة لقيمة "eigen" الثانية تقدم شاهداً جيداً على عاملٍ كامنٍ مهيمن. كما أن هناك دعماً آخر يأتينا للبُعدية الأحادية إذا كانت كل البنود لها ثقل عالي القيمة على العامل الأول (غير المدور).

تجدر الملاحظة أنه ليس من الضروري وجود مجموعة من البنود تكون أحادية البعد بشكل مطلق من أجل التطبيقات العملية لنظرية الإجابة للبند، وإنما بيّن كل من دراسكو وبارسونز 1983 وريشكاسه 1979 وينكر وستاوت 1994 أن عنصراً مهيمناً وحيداً كاف.

تقدير متغيرات البند (Item Parameter Estimation)

يفترض أن تحليل العوامل يعطي عاملاً مهيمناً فريداً نقدر بعد ذلك نواظم البنود. بينت دراسات التشابه مثل (دراسكو 1989 وماكلاخلان ودراسكو 1987) أن التقييم الهامشي لـ "بوك" (بوك وإيكن 1981 وبوك وليبرمان 1970) يجب أن يستخدم هنا. يقدم لنا برنامج الكمبيوتر Bilog (ميزليفي وبوك 1989) خيارين من أجل التقييم الهامشي: الاحتمالية المطلقة (العظمى) واحتمالية Bayesean"بيزيان". تقدم طريقتا التقييم المذكورتان، باستخدام عينات كبيرة، نتائج متشابهة، ولكن مع العينات الصغيرة تكون تقييمات بيزيان أكثر ثباتاً. عادة نقيم متغيرات البند في الحالتين ونأخذ عينات ذات حجم متواضع نسبياً (200-300).

كثيراً ما نشاهد أن واحداً أو أكثر من متغيرات البند المقيمة باحتمالية عظمى تكون متطرفة أكثر مما يمكن تصديقها (مثلاً تباين مقدر يبلغ 3.0) فمثل هذه التقييمات المتطرفة يكون لها عادة أخطاء معيارية عالية جداً.

أنسب طريقة لتفسير متغير تباين الاحتمالية العظمى يبلغ 3.0 ذو خطأ معياري يبلغ 1.2، هو أن عينة أكبر ضرورية من أجل تقييم الاحتمالية العظمى. ولنفترض وجود صعوبة، كالعودة إلى الهند مثلاً لجمع المزيد من المعلومات، نعود إلى تقييم بيزيان تكون تقييمات بيزين عادة أقل تطرفاً، فناظم قدر بـ 3.0 بالاحتمال الأعظم قد يقيّم بـ 1.3 وتكون له أخطاء معيارية أقل. والخطأ المعياري للاحتمالية العظمى الذي يبلغ 1.2 قد يكون 0.20 بهذا التقييم (بيزيان).

الرسوم البيانية الملائمة (Fit Plots)

بعد دراسة نواظم البنود من المهم اختبار المدى الذي يصف فيه نموذجنا من نظرية الإجابة للبند بشكل وافٍ إجابات البند. هناك عدد من الطرق المتوفرة من أجل هذا الهدف. وقد وجدنا أحدها واسمه "الرسوم البيانية الملائمة" ذات فائدة بشكل خاص. فبشكل عام تقارن الرسوم البيانية الملائمة النسبة الفعلية من إجابة المستجيبين لبند ما في فترة θ ثبتا حتى نسبة مقدرة، (أي إلى وظيفة الإجابة للبند) وصعوبة بناء حبكة مناسبة هي أننا لا نعرف قيمة θ ثيتا لكل مجيب. كما أنه من أجل مقياس قصير نتوقع تفاوتات كبيرة نسبياً بين الـ θ ثيتا المقدرة وتلك التي تتبع لكل مجيب.

وصف "دراسكو" و"ليفين" و"تساين" و"وليامز" و"ميد" (1995) أسلوباً يتصدى للمشكلة. فبدلاً من توزيع كل مجيب على زمرة θ وحيدة مبنية على تقرير قابل للنقض، يوزع كل مجيب بشكل متناسب على فئات θ المتعددة بشكل مبني على الاحتمالية الخلفية التي تقع فيها الـ θ ثيتا التابعة لذلك المجيب في تلك المدة و"العدّ غير الحقيقي" لعدد المجيبين الذين يصادف أن يكونوا في تلك المدة للحصول على "نسب إجرائية" تتبنى البنود، هذه النسبة الإجرائية تقارن عندها بوظيفة الإجابة للبند IRF، انظر دراسكو وصحبه من أجل التفاصيل الفنية لهذا الإجراء، وانظر الأوراق التي أصدرها "ستون" (ستون 2003- ستون؛ 2000) من أجل تفاصيل الإجراء المتعلق بها.

مصفوفات التوصيل (linking Metrics):

يستنتج برنامج الكمبيوتر Bilob الافتراض المستنتج أن الخاصية الكامنة لهذا التوزيع الطبيعي المعياري في مجموع السكان الذين أخذت العينة منهم، إلا أنه لا يوجد سبب ملزم مسبقاً يجعلنا نعتقد أن المجموعات السكانية لديها الانحراف المعياري. (أي واحد)، والمعدل ذاته (أي الصفر) لأي خاصية كامنة. ودون فقدان العمومية يمكن أن أو يقاس مجتمع ما بحيث يكون لديه المعدل صفر ووحدة

الانحراف المعياري، ولكن تدرج أو مقياس الخاصية الكامنة لمجتمع آخر يجب أن يوصل بمقياس المجتمع الأول.

اقترحت عدة طرق لوصل مصفوفات الخاصية الكامنة. "قدم لنا (سيغال 1983) عرضاً جيداً" وقد بين البحث (مثلاً سيغال 1983) أن الخط البياني لخواص الاختبار الذي قدمه ستوكنغ ولورد (1983) والذي يصل الأعمال بشكل جيد وبرنامج Equate (الذي قدمه "بيكر" و"القرني" و"الدوسري 1991) يقدمان أداة يمكن استخدامها لأداء هذا النوع من الوصل. يقدم برنامج بيكر وصحبه معاملات انحناء وتقاطع المعاملات (أ و ب على التوالي) لتحول خطي يصل المقاسات. والمعاملات الموصولة هي:

$$\hat{a}^* = \frac{\hat{a}}{A} \; and \; \hat{b}^* = A \times \hat{b} + B.$$

تجدر الملاحظة أن عناصر التباين والتباين المشترك للمصفوفة المرافقة Matrics التغير والتغير المرافق لمؤشرات البند المقدرة يجب أن تحول أيضاً بشكل جيد.

$$Var(\hat{a}^*) = \frac{Var(\hat{a})}{A^2}, Var(\hat{b}^*) = A^2 \times Var(\hat{b}),$$
$$and \; Cov(\hat{a}^*, \hat{b}^*) = Cov(\hat{a}, \hat{b}).$$

ويصبح الوصل عملية تلقائية عندما لا تبدي أية بنود أداءً تفاضلياً عبر الثقافات، إلا أن بعض الأداء التفاضلي للبنود (DTF) يمكن أن يحدث حتى مع أفضل أشكال التكيفات، ويمكن أن يتشوش الوصل عندما تشمل البنود ذات الأداء التفاضلي في التحليل. وبالتالي يمكن استخدام الإجراء التكراري التالي. بعد وصل مبدئي باختبار الأداء التفاضلي للبنود (DTF) من أجل كل بند (توصل الإجراءات لاحقاً)، ثم توضع البنود ذات إحصاءات الأداء التفاضلي للبنود (DTF) جانباً مؤقتاً ويعاد وصل القياسات باستخدام البنود التي ليس لديها أداء تفاضلي للبنود (DTF)

وبعد الوصل تعاد كافة إحصائيات الأداء التفاضلي للبنود ويستمر هذا الإجراء من وصل البنود التي ليس لها أداء تفاضلي DTF وإعادة إحصاء الأداء التفاضلي لكافة البنود إلى أن يتبين للمجموعة ذاتها من البنود حصول أداء، تفاضلي (للبنود) DTF في تكرارين متتاليين. وقد بينت دراسات التماثل (التي أجراها مثلاً كاندل ودراسكو 1988) فعالية الوصل التكراري.

تحليلات أداء البنود التفاضلية متعددة المجموعات:

نظر لورد (1980) وغيره إلى الأداء التفاضلي للبنود Differential ITem Functioning (DIF) من منظور مقارنة مجموعة جوهرية بمجموعة مرجعية. إلا أنه عند توفر المعلومات في عدة مجموعات من أجل التحليل كان أداء اختبارات الدلالة من أجل الاندماجات الزوجية كافة يبدو متوازياً مع استخدام الاختبارات المتعددة من أجل مقارنة معدل المجموعة وليس من أجل تحليل عام للتباين يختبر في الوقت ذاته مساواة المعدلات لكل المجموعات. قام كيم وكوهن وباركن (1995) بمساهمة مهمة لدراسة الأداء التفاضلي للبند DIF، عندما قدموا طريقتهم في تحليل DIF عن طريق مربع Square بين مجموعتين (لورد 1977-1980)، وبشكل محدد، فإن الفرضية الصفرية من أجل اختبار المساواة بين النواظم للبند (i) فيما بين المجموعات هو

H_0: $\mathbf{C}\xi_i = \mathbf{0}$, حيث Matrix C تضادي يحتوي على عدد P من صفوف:

$$\xi = (a_{i1}, b_{i1}, \ldots a_{iK}, b_{iK}),'$$

هو Vector نواظم البنود، وإحصاءات الاختبار التي وضعها كيم وزملاؤه هي:

$$Q_i = (Cv_i)'(C\Sigma_i C')^{-1}(Cv_i),$$

حيث إن Viهو vector يحتوي على تقديرات ناظم البند

$$v_i = \left(\hat{a}_{i1}, \hat{b}_{i1}, \ldots, \hat{a}_{iK}, \hat{b}_{iK}\right),$$

وP هو رتبة C الذي يكون عادة 2 (K - 1) والتوزيع غير المتلاقي Q هو مربع Chi-square بدرجات P من الحرية (كيم وصحبه 1995).

تحليل أداء الاختبار التفاضلي (Differential Test Functioning Analyses)

بالرغم من أنه من النافع جداً تقييم مدى الأداء التفاضلي على مستوى البند، ففي معظم التطبيقات يستخدم الباحثون علامة مقياس كلية وليس بنوداً مستقلة، وعلى ذلك فالأداء التفاضلي للبند DIF كـ DTF بحد ذاته قد لا يكون ذا أهمية في العديد من الظروف وإنما (DTF) أي أداء الاختبار التفاضلي هو الموضوع ذو الأهمية الرئيسة: أي هل يكون لمجيبين لديهم قيمة متساوية في ثيتا ولكن عيناتهم أخذت من ثقافات ويجيبون بلغات مختلفة - درجات كلية متساوية على مقياس ما؟

قدم "راجيو" و"فان درليندن" و"فلير" (1995) طريقة لاختبار (DTF) تحسب حساب آثار البند التعويضية مثل (DTF) الذي يعمل في اتجاه واحد يمكن إلغاؤه بفعل (DIF) على بند آخر يعمل في الاتجاه المعاكس وقيمه طريقة "راجيو" وصحبه في معرض ترجمة مقاييس تستخدم في بحث ما بين الثقافات تكمن في أن البنود التي تبدي (DIF) لا تحتاج حتمياً لأن تزال من المقياس وإنما لا تحتاج البنود للحذف إلا عندما يكون المؤشر (DTF) عالياً وذا دلالة إحصائية. من منظور تحليل (DTF) يمكن لإجراءات (DIF) إزالة بنود ذات قيمة دون حاجة لذلك، فليس هناك حاجة لحذف بنود عندما يكون (DTF) الكلي غير ظاهر.

يمكن استخدام برنامج DFITDUA الذي وضعه "راجيو" وصحبه (1995) لتقييم (DTF) بين أزواج من العينات (كلمة "راجيو" التي ألقاها بذاته في نيسان 2000) يقوم راجيو حالياً بوضع ملحق متعدد المجموعات لتحليل (DTF) وهو يعطي إحصاء كاي مربع Chi Square يستخدم لتقييم أهمية (DTF) بين مجموعتين. عندما يكون مؤشر (DTF) ذا دلالة إحصائية يتعرف برنامج DFITDUA على البنود ذات المساهمة الأعلى لمربع كاي Chi Square ويزيلها.

وجد فلير (1993) أن مؤشر (DTF) كان حساساً بشكل زائد لأحجام العينات الكبرى.

مع هذه الحساسية يوصي راجيو وصحبه (1995) بعدم حذف البنود التي ينتج عنها مربع كاي Chi إلا عندما يكون مؤشر (DTF) الكلي أعلى من 0,006.

ملخص: يظهر أن الإجراء التحليلي الموصوف أعلاه يقدم وسيلة ذات نفع في تقييم التساوي في القياس بالمقاييس المكيفة للاستخدام بين اللغات والثقافات المتعددة، فهو يقيس افتراض بعدية IRT ويدرس مناسبة النموذج المقدر. ثم استخدمت الطرق المدخلة حديثاً التي تقارن تقييمات البنود بين عدة مجموعات وDTF كلية. لتقديم توضيح لهذا الإجراء نقدم فيما يلي اختباراً لتعادل القياسات في أربعة من المقاييس وضعت أصلاً في الولايات المتحدة وترجمت فيما بعد إلى اللغات الإسبانية والبولونية والمراثي ليتم استخدامها في المكسيك وبولونيا والهند على التوالي.

المنهج METHOD

العينات

كجزء من دراسة أكبر تبحث في فعالية الإجراءات المستخدمة بشأن الموارد البشرية فيما بين الثقافات، طبقت دراسة استطلاعية على 939 من الموظفين والعاملين في شركة طباعة ونشر متعددة الجنسيات، طبقت الدراسات على الموظفين في مؤسسات مقرها الولايات المتحدة (عدد أفراد العينة 239 = n)، والمكسيك (عدد أفراد العينة 253 = n) والهند (عدد أفراد العينة 201 = n) وبولندة (عدد أفراد العينة 246 = n) كانت الغالبية العظمى للمجيبين من مستخدمي اللغة الواحدة. وقد أخذت الغالبية العظمى في عينات المجيبين من الإدارة والعاملين في الإنتاج ومختلف المستويات الإدارية حسب التسلسل الهرمي. كما أبلغ المجيبون بأن إجاباتهم ستكون سرية بشكل مطلق وأن مساهمتهم تطوعية بحتة بإرادتهم فقط. كما أنهم أعلموا بأن إدارات مؤسساتهم والمستويات العليا من الإدارة في كل مؤسسة قد أعطت إذناً بإجراء هذه الدراسة ضمن أوقات العمل.

المقاييس:

قيم الاستبيان خواص المجيبين الديمغرافية (السكانية) وبيئتهم الأساسية (الخلفية)، كما أن البحث اشتمل على مقاييس ذات تدريجات لتقيس أنماط السلوك القيادية الموحية بالقوة وفرص التقدم المستمر والاستقلالية (العزلة) التنظيمية والالتزام التنظيمي.

وقد كانت المقاييس المستخدمة لقياس الجوانب المختلفة من الرضى عن العمل والضغط العام للوظيفة من الأمور التي أولياها الاهتمام الخاص في هذا الفصل. وقد استخدمت نسخ ذات تسعة بنود من المقاييس الفرعية: الرضى عن العمل ذاته، والرضى بالإشراف والرضى بالعاملين مع الشخص المجيب، وهي مأخوذة من دليل توصيف العمل (JDI) بقلم ("سميث" و"كيندال" و"هونين" 1969) منقحة من قبل ("رزونوفسكي" 1989). وقد استعمل المشاركون المجيبون مقياس إجابة مدرج بثلاث درجات من "نعم" أو "لا" أو "لا أعرف" ليبينوا مدى وصف أي من الصفات أو العبارات لوظيفتهم. يحتوي المقياس العام (الذي وضعه "سيدمان" و"ماكراري" عام 1992) على صيغة إجابة تماثل صيغة (JDI) ويقيس التوتر الوظيفي العام. يحتوي الجدول 11-1 على المقاييس والبنود التابعة لها بأكملها.

التحليلات

تسجيل المتغيرات:

سجلت الإجابات لمقياس (JDI) في علامات ثنائية كما هو مطلوب من أجل نموذج (IRT) المستخدم في هذه الدراسة. وقد أ عطيت درجة واحدة للبنود الإيجابية المجاب عليها بالإيجاب والبنود السلبية المجاب عليها بالنفي. أما الإيجاب بالنسبة للبنود السلبية وعدم الإيجاب مع البنود الإيجابية فقد كان الرمز المعطى لها صفراً. أما الإجابات المقابلة لإشارة الاستفهام فقد أعطيت علامة الصفر نتيجة لما وجد إجرائياً في أن مثل هذه الإجابات كانت مرتبطة بشدة بعدم الرضى بالوظيفة

أكثر من ارتباطها بالرضى بالوظيفة. (هانيش 1992- سميث 1969) وقد وضعت الدرجات بالنسبة لمقاييس "SIG" بشكل موازٍ، أما الإجابات التي تشير إلى توتر أعلى فكانت تعطى علامة /1/، والإجابات التي تبيّن قدراً أقل من التوتر تعطى علامة صفر. إلا أن إجابات إشارة الاستفهام فقد أعطيت علامة /1/. كما أن مؤشرات إجابات المقياس كافة التي تحتوي على نقص لأكثر من بند واحد قد استبعدت من التحليلات التالية فيما بعد.

الجدول رقم 11-1

بنود الرضى بالمشاركين في العمل وبالمشرف عليه وبالعمل ذاته من فهرس توصيف وظيفي ومقياس التوتر بشكل عام

	United States				*Poland*				*Mexico*				*India*			
	p	r_b	*a*	*b*	*p*	r_b	*a*	*b*	*p*	r_b	*a*	*b*	*p*	r_b	*a*	*b*
Coworker Satisfaction																
Boring	.840	.624	.922	−1.448	.854	.878	.241	−1.350	.923	.694	.907	−2.112	.796	.646	.948	−1.212
Slow	.690	.773	1.179	−.638	.758	.643	.890	−1.102	.774	.788	1.161	−.951	.773	.929	1.366	−.937
Loyal	.426	.636	1.084	.287	.536	.679	1.118	−.101	.557	.678	1.219	−.128	.617	.661	.969	−.409
Responsible	.706	.739	1.134	−.702	.686	.728	1.087	−.647	.812	.967	1.396	−1.046	.759	.819	1.156	−.927
Waste of time	.651	.752	1.170	−.498	.808	.794	1.108	−1.214	.484	.258	.593	.118	.630	.426	.729	−.545
Lazy	.634	.792	1.287	−.420	.741	.828	1.189	−.887	.759	.752	1.121	−.916	.790	1.005	1.465	−1.000
Unpleasant	.758	.648	.962	−.976	.845	.809	1.132	−1.422	.844	.786	1.058	−1.348	.693	.784	1.135	−.650
Intelligent	.664	.686	1.017	−.563	.683	.811	1.264	−.605	.747	.662	.971	−.898	.716	.336	1.031	−.774
Work well together	.664	.686	1.024	−.567	.716	.782	1.178	−.759	.780	.855	1.241	−.963	.819	.761	1.066	−1.241

(continued on next page)

TABLE 11.1 (continued)

	United States				*Poland*				*Mexico*				*India*			
	p	r_b	*a*	*b*	*p*	r_b	*a*	*b*	*p*	r_b	*a*	*b*	*p*	r_b	*a*	*b*
Supervisor Satisfaction																
Hard to please	.515	.770	1.198	−.027	.379	.539	.865	.525	.648	.726	1.171	−.446	.286	.304	.689	.958
Impolite	.690	.782	1.178	−.660	.780	.881	1.308	−1.302	.855	.992	1.413	−1.289	.534	.793	1.356	−.140
Praises good work	.414	.638	.969	.338	.610	.577	.821	−.431	.728	.514	.806	−.930	.609	.727	1.122	−.358
Tactful	.450	.479	.729	.231	.599	.748	1.080	−.317	.636	.519	.817	−.505	.441	.061	.508	.239
Annoying	.573	.816	1.289	−.222	.640	.873	1.351	−.484	.800	.915	1.257	−1.052	.537	.800	1.377	−.136
Bad	.715	.875	1.426	−.713	.751	.890	1.320	−.936	.861	.900	1.292	−1.311	.637	.771	1.291	−.491
Interferes with my work	.745	.775	1.176	−.882	.785	.803	1.141	−1.109	.588	.291	.592	−.372	.525	.775	1.233	−.107
Gives confusing directions	.603	.668	.965	−.365	.662	.863	1.294	−.552	.714	.761	1.164	−.676	.576	.772	1.202	−.268
Knows how to supervise	.477	.813	1.294	.100	.580	.934	1.575	−.247	.750	.754	1.168	−.851	.551	.656	.944	−.187

Work Satisfaction																
Fascinating	.303	.759	1.288	.721	.203	.762	1.226	1.079	.604	.569	.915	−.353	.553	.511	.832	−.234
Satisfying	.601	.865	1.343	−.340	.543	.709	1.119	−.167	.783	.789	1.209	−.989	.822	.952	1.401	−1.183
Boring	.657	.729	.999	−.597	.648	.784	1.328	−.512	.849	.667	.982	−1.437	.777	.713	1.001	−1.152
Creative	.422	.773	1.151	.294	.244	.660	1.032	.941	.744	.600	.954	−.879	.559	.599	.942	−.262
Challenging	.686	.719	.996	−.721	.823	.413	.758	−1.564	.550	.311	.649	−.209	.697	.808	1.176	−.720
Gives sense of accomplish-ment	.644	.924	1.518	−.478	.404	.737	1.134	.319	.838	.814	1.219	−1.235	.731	.932	1.441	−.767
A source of pleasure	.343	.786	1.315	.570	.245	.733	1.179	.912	.500	.552	.999	.026	.620	.485	.789	−.506
Dull	.665	.846	1.233	−.586	.438	.642	.985	.221	.500	.589	1.000	.002	.620	.485	.789	−.506
Interesting	.628	.966	1.678	−.409	.534	.841	1.412	−.118	.846	.862	1.252	−1.271	.680	.730	1.058	−.723

(continued on next page)

TABLE 11.1 (continued)

	United States				*Poland*				*Mexico*				*India*			
	p	r_b	*a*	*b*	*p*	r_b	*a*	*b*	*p*	r_b	*a*	*b*	*p*	r_b	*a*	*b*
Stress in General																
Hectic	.612	.682	.975	−.432	.409	.527	.910	.282	.576	.452	.850	−.306	.589	.644	1.022	−.364
Tense	.565	.850	1.371	−.218	.641	.533	.894	−.566	.282	.696	1.131	742	.481	.794	1.216	.032
Frantic	.278	.821	1.367	.788	.298	460	.869	.701	070	534	.895	2.091	.350	.829	1.294	.507
Pressured	.648	.774	1.186	−.523	.813	.018	.530	−1.841	.372	.753	1.358	.381	.508	.841	1.313	−.051
Hassled	.333	.727	1.092	.616	.495	.572	.955	−.036	.080	.586	.917	2.000	.406	.882	1.409	.287
Relaxed	.638	.767	1.164	−.494	.840	230	.680	−1.1661	.344	628	1.138	.531	.520	.256	.574	−.154
Many things stressful	.612	.811	1.256	−.394	583	.682	1.179	−.308	.448	.510	.858	.169	.500	.860	1.346	−.024
Nerve-wracking	.381	.845	1.390	.400	.473	.656	1.141	−.051	.162	.739	1.089	1.292	.307	.809	1.238	.649
More stressful than I'd like	.475	.744	1.075	.095	.378	.643	1.109	.327	.284	.304	.660	.979	.369	.778	1.048	.455

اختبار أحادية الأبعاد

من أجل اختبار مناسبة افتراض أحادية الأبعاد بالنسبة للمقاييس الأربعة أجريت تحليلات للعوامل الأساسية (الرئيسة) تكرارياً وذلك في كل بلد كما وصفنا سابقاً. وقد استخلصت العوامل ذات قيمة eigen التي تكون أعلى من 1.0 وإذا حصلنا على استخلاصات وحيدة العامل أو شحنات عالية على عامل المحور الرئيس الأول (PAF) فذلك يعني أن مقياس "ai" يقيس تركيبة مهيمنة واحدة وذلك يدعم الافتراض بأحادية البعد.

تقييم متغيرات البنود Item Parameter Estimates

استخدم بيلوغ "Bilog" (مليفي وبوك 1989) لتقييم نواظم البنود للنموذج التنظيمي ذي الناظمين بالنسبة لكل مقياس في كل من العينات الأربع. حدد المدى الأقصى لدورات الـ(EM) المئة ولدورات "نيوتن -رافسون" العشر وقد استخدمت 30 نقطة تربيعية بدلاً عن العدد الناقص وحدد مبدأ الدمج ذو قيمة 0.001 وحدد توزيع مسبق طبيعي N (0.1) من أجل ناظم صعوبات البنود. بالإضافة إلى ذلك حدد توزيع مسبق طبيعي جذعي، N (0,02) من أجل نواظم التفريق بين البنود.

وصل القياسات تكرارياً

استخدم برنامج بيكرو (1991) "EQUATE" لوصل قياسات كل مجموعة مركزية (المكسيك، وبولندة والهند) إلى مجموعة المرجع (الولايات المتحدة) بطريقة ستوكينغ -لورد (1983)، كما استخدم الإجراء التكراري الذي وصفه كاندل ودراسكو (1988) وليونغ ودراسكو (1986) في إجراء الوصل، وهنا استخرجت مجموعة مبدئية لمعاملات الوصل لتحويل تقييمات النواظم من كل في المجموعات المركزية (موضوع الدراسة) إلى قياسات مجموعة المرجع. أجري تحليل (DIF) متعدد المجموعات لـ (كيم وصحبه 1995) واستبعدت البنود التي وجد أنها تبدي

(DIF) في مجموعة البنود المدروسة وطبقت طريقة "ستوكينغ -لورد" في الوصل ثانية باستخدام البنود غير المتأثرة بالانحراف فقط؛ وبعد إعادة الوصل أعيد اختبار كافة البنود من أجل (DIF) وتم تكرار هذا الأسلوب إلى أن ظهرت المجموعة ذاتها من البنود (DIF) في محاولتين متتاليتين أجري تحليل (DTF) بعد وصل تقييمات البنود في المجموعات الثلاث موضوع الدراسة بمجموعة المرجع بطريقة الإجراء التكراري الموصوف سابقاً. وهنا قورنت كل من المجموعات الثلاث موضوع الدراسة (المكسيك، بولندة، الهند) بالمجموعة المرجع (الولايات المتحدة) لتقرير ما إذا كان هناك وجود لأداء المقياس التفاضلي الكلي لكل من المقاييس الأربعة.

النتائج

أوردنا إحصائيات نظرية الاختبار التقليدية، أي نسبة "الصحيح" وارتباطات مجموع البنود ذات السلسلة الثنائية من أجل التوتر في المقياس العام ومن أجل مقاييس الرضى بالعمل والرضى بالمشرف والرضى بالعاملين المرافقين في الجدول رقم 11-1 .

البُعدية Dimensionality

بينت التحليلات العاملية المجراة في كل عينة من أجل كل مقياس بأن المقاييس كانت أحادية الأبعاد بما فيه الكفاية للشروع بتحليلات (IRT) وقد بيّنا نتائج (PAF) في الجدول 11-2 .

الرضى بزملاء العمل المرافقين

نتج عن تحليل العاملين لمقياس الرضى بزملاء العمل استخراج عامل وحيد في كل عينة. وقد فسر هذا العامل أكثر من ثلث التباين في كل عينة. وقد كان حمل البنود عالياً بعض الشيء في الولايات المتحدة يتراوح بين /0.47/ (ممل) إلى /0.68/ (كسول) (متراخ)، وتباينت الأحمال في المكسيك بين /0.43/ (ممل) و/0.75/ (يحمل المسؤولية) وتراوحت الأحمال في "الهند" من /0.36/ (تضييع

للوقت) إلى /0.83/ (كسول) وأخيراً في بولندة كانت الأحمال أيضاً عالية تراوحت بين /0.51/ (بطيء) و/0.67/ (ذكي).

الرضى بالمشرف على العمل:

نتج التحليل العاملي لمقياس الرضى بالمشرف على العمل المكون من تسعة بنود عن استخلاص عامل وحيد أيضاً في كل عينة مما فسر تبايناً يزيد عن الثلث في كل عينة. وقد كانت حمولات كل بند عالية شيئاً ما تتراوح بين /0.41/ (لبق) إلى /0.73/ (سيئ) في الولايات المتحدة. وهي المكسيك بين /0.25/ (يتدخل في عملي) إلى /0.71/ (يعرف كيف يشرف على الناس) وفي الهند بين /0.08/ (لبق) إلى /0.78/ (مزعج) وفي بولندة /0.46/ (صعب إرضاؤه) إلى (يعرف كيف يشرف على الناس).

الرضى بالعمل ذاته:

أدى التحليل العاملي لمقياس الرضى بالعمل، المكون من تسعة بنود إلى استخلاص عامل وحيد في كل من العينات الأربع. وقد تراوح التباين بموجب هذا العامل بين 25.5٪ في المكسيك إلى 38.7٪ في الولايات المتحدة. وقد كانت الحمولات الكلية في كل عينة عالية بعض الشيء، فقد تراوحت بين /0.16/ (ممل) و/0.80/ (مشوق) (يدعو للاهتمام) في الولايات المتحدة. وفي المكسيك تفاوتت الأحمال بين /0.14/ (ممل) و/0.77/ (مشوق يدعو للاهتمام) وفي الهند تروحت الأحمال بين /0.31/ (ممل) و/0.76/ (مرضي)، وأخيراً في بولندة تراوحت الأحمال بين /0.06/ (ممل) و/0.72/ (مشوق). ومن الممتع ملاحظة الميل للتقارب بين الحمولات لعوامل بين العينات، ففي كل دولة كانت البنود (ممل) و(كئيب) حمولة منخفضة بشكل متناسق بينما كان البند (مشوق) دائماً حمولة عالية على العامل العام.

الضغط العام للعمل (التوتر):

نتج عن تحليل PAF لمقياس "التوتر بشكل عام" المكون من تسعة بنود (تدريجات) أيضاً عامل وحيد استخلص من كل عينة، وقد تراوحت التباينات المقاسة من 25.7٪ في المكسيك إلى/0.45٪ في الهند وكانت الحمولات العاملية عالية في كل من العينات في الولايات المتحدة تراوحت الحمولات من 0.56 مضطرب إلى 0.79 (متوتر).

الجدول رقم 11-2

نتائج تحليلات للعاملية للمحور الرئيس من أجل الرضى بالمشاركين في العمل وبالمشرف وبالعمل ذاته ومعدلات التوتر بشكل عام بحسب الدولة.

	United States		*Poland*		*Mexico*		*India*	
	# factors	*% variance*	*# factors*	*% variance*	*# factors*	*% variance*	*# factors*	*% variance*
:ale								
oworker	1	35.9	1	37.9	1	35 2	1	37.9
upervisor	1	40.0	1	42.8	1	33 8	1	37.6
ork	1	38.7	1	29.4	1	25 5	1	33.0
ob Stress	1	42.8	1	26.3	1	25.7	1	45.0

وتراوحت الحمولات في المكسيك بين /0.28/ (مضطرب) و/0.61/ (محطم للأعصاب) وفي الهند تراوحت الحمولات بين /1.19/ (مريح) و/0.77/ (متوتر) وأخيراً في بولندة كانت الحمولات تتراوح بين /0.13/ (مريح) و/0.63/ (محطم للأعصاب) فهنا أيضاً كانت هناك درجة من الاتساق بين العينات فيما يتعلق بالحمولات العالية والحمولات المنخفضة للبنود.

تقييمات ناظم البند:

وضعت الإجابات المعطاة علامات ثنائية والمستخلصة في 939 موظفاً كما وصفنا سابقاً كمدخل في Bilog وكان النموذج الدلالي قيد قيّم من أجل كل مقياس

باستخدام المعلومات المأخوذة من كل قطر. وقد وضعت تقييمات النواظم الناتجة في الجدول 11-1 وجرت مناقشتها فيما يلي:

الرضى بزملاء العمل

بينت القياسات المبدئية (أي قبل التعادل) Bilog للشعور بالرضى عن زملاء العمل في كل من الدول الأربع أن البنود كانت بشكل عام (سهلة) (أي أدخلت إيجابياً في الكثير من المرات) وكانت كافة التقييمات (فيما عدا اثنين) ويبلغ تعدادها 36 ذات تقييمات سلبية. وهكذا فالأفراد ذوو مستويات الرضى المتراوحة بين نصف وواحد انحراف معياري (SD) تحت المعدل كان لديهم احتمال بنسبة 50٪ الادخال العديد من البنود إيجابياً (أي تبيان الرضى). كان معدل تقييم في الولايات المتحدة - 614.- (SD = 0.416) وكان يتراوح بين (1.448 - ممل) إلى (0.287 مُوَالٍ) وكان معدل تقييم الصعوبة في المكسيك 916.- (SD = 0.645). وفي الهند، كان معدل التقييم 0.844 (SD = 0.42). أما تقييم النواظم فقد تراوح بين 1.241-(يعمل جيداً مع الآخرين) إلى 409.- (مُوَالٍ). أخيراً في بولندة كان المعدل 899.- (SD = 0.422) بالإضافة لذلك بدا أن البنود تميز بشكل جيد.

وكان معدل تقييم التمييز لبنود المشاركين في العمل في الولايات المتحدة 1.086 (SD = 0.117) ويتراوح بين 0.922 (ممل) إلى 1.287 (كسول). وكان المعدل في المكسيك لتقييم ai هو 1.074 (SD = 0.233)، وتراوحت التقييمات بين 0.593 (تضييع للوقت) إلى 1.396 (شعور بالمسؤولية) وكان معدل التقييم في الهند 1.096 (SD = 0.222) وكانت التقييمات تتراوح بين 0.729 (هدر للوقت) إلى 1.465 (كسول) أخيراً في بولندة كان معدل تقييم ai 1.134 (SD = 0.109) ويتراوح بين 0.890 (بطيء) إلى 1.264 (ذكي)، وهذه المعدلات تبين أن البنود تتمايز بشكل جيد.

الرضى بالمشرف على العمل:

كانت بنود الرضى بالمشرف بشكل عام (سهلة) أيضاً بالرغم من أن معدل تقييمات bi كان أكبر منه بالنسبة لبند الرضى بزملاء العمل. كان معدل صعوبة

البنود في الولايات المتحدة 244.- (SD= 0.439) وتراوحت التقديرات بين 882.- (يتداخل بعملي) إلى 0.338 (يمتدح العمل الجيد) وقد كان معدل تقييم b_i في بولندة 0.509 (SD= 500.0) وتراوحت تقييمات الناظم بين 1.302. (غير مهذب) إلى 0.525 (من الصعب إرضاؤه)، وفي المكسيك كان معدل الصعوبة 0.826- (SD= 0.351) وتراوحت التقييمات بين 1.311- (سيئ) إلى 0.372- (يتدخل في عملي)، أخيراً في الهند كان معدل لتقييم 0.054- (SD= 0.429) وتراوحت تقييمات الناظم بين 0.491- (سيئ) و0.958 (من الصعب إرضاؤه).

وباختبار تقييمات a_i نجد أن بنود الرضى بالمشرف تميز جيداً باستثناءات قليلة جداً في الولايات المتحدة كان تقييم معدل التمييز 1.136 (SD= 0.213) وتتراوح التقييمات بين 0.729 (لبق) و1.426 (سيئ). في بولندة كان معدل تقييم a_i 1.195 (SD=0.243) وكانت التقييمات تتراوح بين 0.821 (يمتدح العمل الجيد) و1.575 (يعرف كي يشرف على الناس) وكان معدل التمييز بين البنود في المكسيك 1.075 (SD= 0.272) وتراوحت التقييمات بين 0.592 (يتداخل بعملي) و1.413 (غير مهذب) أخيراً كان معدل تقييم a_i في الهند 1.080 (SD= 0.306) وتراوحت التقييمات بين 0.508 (لبق) و1.377 (مزعج).

الرضى بالعمل:

تظهر دراسة تقييمات ناظم صعوبة البنود أن هناك معدلاً واسعاً من صعوبة البنود في مجال قياس الرضى بالعمل وبالذات في الولايات المتحدة وبولندة، وتراوحت تقييمات صعوبة البنود في الهند والمكسيك على العموم في نهاية المجال المرقم بـ (سهل).

كان معدل صعوبة البند في الولايات المتحدة 172.- (SD = 0.547) وتراوحت التقييمات بين 721.- (يمثل تحدياً) و0.721 (ذو سحر) وكان معدل تقييمات b_i في بولندة 0.123 (SD= 0.840) وتتراوح التقييمات بين 1.564- (يمثل تحدياً) و1.079

(ذو سحر) وفي المكسيك كان معدل تقييم bi 705.- (SD = 0.576) حيث تتراوح التقييمات بين 1.437- (ممل) و0.026 (مصدر للسرور). أخيراً في الهند كان معدل صعوبة البند 697.- وتتراوح تقييمات ناظم البنود بين 1.183- (مرضي) و234.- (ذو سحر) في العينات الأربع. وقد كان معدل تمييز البند في الولايات المتحدة 1.280 (SD = 0.223) بتمييز للتقييمات يتراوح بين 0.996 (يمثل تحدياً) و1.678 (مشوق). وفي بولندة كان معدل تقييم ai 1.130 (SD = 0.194)، وكان تقييم ناظم البنود يتراوح بين 0.758 (يمثل تحدياً) و1.328 (ممل). وكان معدل تقييم التمييز في المكسيك 1.020 (SD = 0.189) حيث تتراوح تقييمات البند بين 649.- (يمثل تحدياً) و1.252 (مشوق). أخيراً في الهند كان معدل تقييم المنحى 1.113- (SD= 0.248) وتتراوح التقييمات بين 0.832 (ذو سحر) و1.441 (يعطي شعوراً بالإنجاز).

التوتر العام في الوظيفة:

أظهرت معايرة Bilog لمقياس التوتر بشكل عام المكون من 9 بنود انتشاراً جيداً لصعوبة البنود. كان معدل الصعوبة في الولايات المتحدة 0.018 (SD = 0.509) وتتراوح تقييمات ناظم البنود بين 523.- (عليه ضغط) و0.788 (يحطم الأعصاب). وكان معدل صعوبة البند 339.- (SD = 0.811) وكانت صعوبات البنود تتراوح بين 1.841- (عليه ضغط) و0.701 (يحطم الأعصاب). وفي المكسيك كان معدل تقييم bi 0.876 (SD =0.807) وتتراوح التقييمات بين 0.306- (مقلقل) إلى 2.091 (يدعو للجنون). أخيراً في الهند كان المعامل 0.149 (SD = 0.341) وتتراوح تقييمات الصعوبة بين 394.- (مقلقل) و0.649 (محطم للأعصاب).

وقد كانت تقييمات تمييز البنود بشكل عام جيدة جداً؛ فقد كان معدل تقييم ai في الولايات المتحدة 1.208 (SD = 0.148) والتقييمات تتراوح بين 0.975 (مقلقل) و1.390 (محطم للأعصاب). وفي بولندة كان معدل التقييم 0.919 (SD = 0.214)، وتتراوح التقييمات بين 0.530 (تحت ضغط) و1.179 (كثير من الأشياء متوترة)،

وفي المكسيك كان معدل تقييم ai 0.988 (SD = 0.209)، والتقييمات تتراوح بين 0.660 (أكثر توتراً مما أحب) و1.358 (تحت ضغط). أخيراً كان معدل التقييم في الهند 1.162 (SD = 0.256) والتقييمات تتراوح بين 0.574 (مريح) و1.409 (ملاحق).

وصل القياسات مع أداء البند التفاضلي متعدد المجموعات:

حيث إن تقييمات الناظم البنود الموصوف سابقاً عشوائية فيما يتعلق بمصدر الوحدة فقد وصلت قياسات كل عينة مركزية (موضوع الدراسة) للمجموعة بمجموعة المرجع باستخدام أسلوب الوصل التكراري الموصوف سابقاً. اختيرت الولايات المتحدة كمجموعة مرجعية لأن لغة المصدر لكل من المقاييس المستخدمة المكيفة هي اللغة الانكليزية. إلا أنه من المهم التنويه بأن أياً من هذه المجموعات موضوع الدراسة يمكن أن تختار كمجموعة مرجعية بالدرجة نفسها تقارب الوصل في تكرارين لكل عينة وكل مقياس. يمكن رؤية ثوابت التحويل في كل تحويل في الجدول 11-3. ونتبنى مقياس الرضى بالمشاركين في العمل لتوضيح إجراء الوصل التكراري. بعد الوصل المبدئي للمجموعات موضوع الدراسة بالمجموعة المرجعية (مثلاً A = 0.943 و B = 0.279 بالنسبة للمكسيك). طبق تحليل DIF متعدد المجموعات وجد DIF في بند واحد (هدر الوقت) لذلك استبعد هذا البند من المقياس وأعيد حساب معاملات الوصل باستخدام البنود التسعة الباقية. طبقت معاملات الوصل الجديدة (مثلاً: A = 1.037 و B = 0.446 بالنسبة للمكسيك) وأعيد حساب إحصاءات DIF فوجد أن بند (هدر الوقت) ما زال يبدي DIF وحيث أنه لا يوجد DIF بالنسبة لأي بند آخر انتهى إجراء الوصل التكراري.

الجدول3.11

معاملات التحولات الخطية التي تصل المجموعات موضوع الدراسة بالمجموعة المرجعية من أجل كل بعد للوصف الوظيفي والتوتر الوظيفي

	Iteration 1		*Iteration 2*	
	A	*B*	*A*	*B*
Coworker Satisfaction				
Mexico	943	.279	1.037	446
India	1.012	.272	1.081	.352
Poland	1.044	.334	1.059	291
Supervisor Satisfaction				
Mexico	.941	.540	1.015	.727
India	.932	-.170	.907	-.094
Poland	1.039	.288	1.046	.297
Work Satisfaction				
Mexico	.808	.422	841	712
India	.905	.478	902	.627
Poland	.851	- 305	905	-.346
Job Stress				
Mexico	.779	-.687	757	- 706
India	.971	-.172	1.062	- 167
Poland	.731	.176	.746	.232

من المهم جداً تقدير قيمة كاي مربع Chi الحاسمة التي ستستخدم في تصنيف البنود على أنها تبدي DIF أو لا تبديها. يمكن لهذه القيمة الحاسمة أن تعتمد على عدد المجموعات المركزية (موضوع الدراسة) وعدد النواظم في نموذج IRT والنوع المرغوب فيه من مستوى Alpha وعدد البنود في المقياس. في هذه الدراسة كانت هناك ثلاث مجموعات موضوع الدراسة وناظمين وعلى ذلك فقد كان لإحصاء اختبار كاي مربع Chi ست درجات من الحرية إذا رغبنا في إبقاء Alpha كلية قيمتها 0.01 من أجل اختبار DIF على كل مقياس فإن تطبيق تصحيح بوفروني

سينتج ألفا Alpha لكل بند قيمتها 0.001 تقريباً ومربع كاي Chi حاسم قيمته 22.46 ومما تجدر ملاحظته أن كثيراً من البنود كان لها قيم مربع Chi تقرب من هذه القيمة (انظر الجدول 11-4 من أجل قيم مربع كاي Chi لكل بند من التكرارات 1 و2). وقد بينت حبكات IRF لمثل هذه البنود اختلافات صغيرة نسبياً، وعلى ذلك قررنا تصنيف البنود ذات / $\chi^2 \geq 50$ / على أنها قيم تبدي DIF تُظْهِر قِيمُ الـ (IRF) لمثل هذه البنود اختلافات عالية نسبياً). وبالنسبة لكل من المقاييس الثلاثة الأخرى (الرضى بالإشراف، الرضى بالعمل ذاته، التوتر بشكل عام) فإن عملية الوصل كررت مرتين فقط. وفي كل حالة وجد أن البنود المتعارف عليها على أنها منحازة في المرة الأولى كانت مماثلة للبنود التي وجد أنها منحازة في التكرار الثاني.

وكانت البنود التي أبدت (DIF) في المقاييس الأربعة قليلة نسبياً. وكان هناك بند واحد من مقياس "الرضى بالمشاركين في العمل" وجد أنه منحاز وهو "هدر الوقت" وكان له قيمة $\chi^2_{(6)}$ = 78.90 وهناك بند واحد من مقياس الرضى بالمشرف وُجد أنه يبدي () وهو: يتدخل في عملي وكان له $\chi^2_{(6)}$ = 93.23 وهناك بندان على مقياس الرضى قد أظهرا DIF وهما: «التحدي» وكانت له قيمة $\chi^2_{(6)}$ = 141.23 «الملل» وكانت له قيمة $\chi^2_{(6)}$ = 086.48 . أخيراً وجد بندان من مقياس التوتر بشكل عام منحازين: (مقلق) الذي كان له $\chi^2_{(6)}$ =111.43 و(مريح) بقيمة $\chi^2_{(6)}$ = 63.05.

ومن المهم أن نتذكر أن تحليل (DIF) متعدد المجموعات لا يقرر فقط أن هناك (DIF) بين مجموعتين أو أكثر من المجموعات تحت الدراسة. إن هذا التحليل لا يحدد أين ينشأ الأداء التفاضلي. ومن أجل تقرير ذلك يجب إجراء مقارنات ثنائية باستخدام مربع كاي Chi الذي قال به لورد.

تحليل أداء الاختبار التفاضلي:

من أجل تقييم الأداء التفاضلي على مستوى الاختبار، أجريت مقارنات ثنائية بين عينة الولايات المتحدة كمجموعة مرجعية، وكل من العيّنات الثلاث الأخرى موضوع الدرس. وبعد إجراء تقرير مستوى (DTF) بيّن برنامج DFITDUA البنود

التي يستحسن إزالتها للقضاء على الأداء التفاضلي الكلي للاختبار. وكان البند الذي يجري اختياره للاستبعاد هو الذي يساهم بشكل ذي دلالة في الـ DTF الكلي ويتم استبقاؤه إذا كان يؤدي إلى مؤشر DTF قيمته 0.006 أو أعلى (راجيو وصحبه 1995).

الجدول 11-4

احصاءات مربع كاي Chi للأداء التفاضلي للبند متعدد المجموعات للتكرارين 1 و 2

	Iteration 1 χ^2	*Iteration 2* χ^2
Coworker Satisfaction		
Boring	11.75	10.42
Slow	6.09	5.64
Loyal	6.46	5.46
Responsible	10.36	5.98
Waste of Time	63.50*	78.90**
Lazy	4.41	4.79
Unpleasant	16.78	17.92
Intelligent	3.18	2.56
Work Well Together	8.03	5.02
Supervisor Satisfaction		
Hard to Please	39.30	35.71
Impolite	10.88	13.45
Praises Good Work	34.10	30.51
Tactful	15.95	13.43
Annoying	3.20	1.27
Bad	0.65	1.90
Interferes with my Work	76.04*	93.23**
Gives Confusing Directions	6.95	9.85
Knows How to Supervise	12.39	7.48
Work Satisfaction		
Fascinating	25.98	10.59
Satisfying	3.99	7.94

	Iteration 1 χ^2	*Iteration 2* χ^2
Boring	7.40	10.33
Creative	31.02	12.41
Challenging	105.16*	141.23**
Gives Sense of Accomplishment	20.13	11.72
A Source of Pleasure	23.52	28.12
Dull	51.91*	86.48**
Interesting	16.53	13.84
Job Stress		
Hectic	51.53*	111.43**
Tense	4.16	2.57
Frantic	18.05	7.90
Pressured	16.91	22.58
Hassled	22.08	17.32
Relaxed	64.20*	63.05**
Many Things Stressful	28.84	23.68
Nerve-Wracking	1.12	3.64
More Stressful Than I'd Like	34.45	23.22

الرضى بالمشاركين في العمل:

لقد أظهرت المقارنة بين الولايات المتحدة والمكسيك أدائية اختبارية ذات دلالة: DTF = 0.034، $X^2(252) = 318.89$، $p < 0.01$. وقد تم اقتراح استبعاد أربعة بنود وهي «الملل»، «هدر الوقت»، «الكسل»، «العمل الجماعي الجيد». وعند استبعاد هذه البنود فإنه يمكن الحصول على قيمة دليل DTF = 0.004 و $X^2(252) = 705.81$، و$p < 0.001$ ، وتجدر الإشارة إلى أن قيمة مربع كاي Chi أصبحت أكبر وذات قيمة إحصائية واضحة؛ وفي الوقت نفسه فإن قيمة دليل DTF أصبحت أقل من 0.006 وطبقاً لبحث فلير (1993) فإنه يجب عدم استبعاد أي من البنود الأخرى.

وأظهرت المقارنات بين الولايات المتحدة والهند أدائية اختبارية تفاضلية ذات دلالة: DTF = 0.011، وX^2(200) = 958.09، وp < 0.001 وقد تم اقتراح استبعاد بند واحد وهو «هدرالوقت» وقد نتج عن استبعاد هذا البند مؤشر DTF = 0.002، وX^2(200) = 208.88.

وبالمقارنة بين الولايات المتحدة الأمريكية وبولندا ظهر أن هناك أدائية اختبارية تفاضلية ذات دلالة: DTF = 0.015، X^2(245) = 462.16، p < 0.001 وتم اقتراح استبعاد بند واحد وهو «هدر الوقت». وباستبعاد هذا البند فقد تم الحصول على مؤشر DTF = 0.000، وهكذا لم يتم استبعاد أية بنود أخرى.

الرضى بالإشراف على العمل:

بينت المقارنات بين الولايات المتحدة والمكسيك أدائية اختبارية تفاضلية ذات دلالة:

DTF= 0.085 ، X^2 (252) = 1635.39، p <0.001 وتم اقتراح استبعاد بند واحد وهو (يتدخل بعملي) ونتج عن استبعاد هذا البند مؤشر DTF قيمته 0.000.

وبينت المقارنات بين الولايات المتحدة والهند أدائية اختبارية تفاضلية ذات دلالة:

DTF = 0.055 ، X^2 (200) = 609.07، p <0.001 وتم استبعاد بند واحد هو (يتدخل بعملي) ونتج عن حذف هذا البند مؤشر DTF قيمته 0.009 وX^2 (200) = 200.11 وعلى ذلك لم تحذف أية بنود أخرى.

ولم ينتج عن مقارنة الولايات المتحدة ببولندة أدائية اختبارية تفاضلية ذات دلالة، DTF = 0.002 وعلى ذلك لم تحذف أية بنود أخرى.

الرضى بالعمل نفسه:

بينت المقارنات بين الولايات المتحدة والمكسيك فروقات جوهرية فكانت القيم:

DTF = 0.466 , X^2(252) = 1568.25 ، P < 0.001 وتم اقتراح استبعاد بندين

هما: «التحدي» و«الملل». وباستبعاد هذين البندين تم التخلص من الفروقات (مؤشر DTF = 0.004).

وبينت المقارنات بين الولايات المتحدة والهند أيضاً اختلافات جسيمة فكانت القيم: DTF = 0.164، $X^2(200)$ = 518.94، P<0.001 . وتم اقتراح استبعاد بندين وهما: «التحدي و «الملل». ونتيجة لحذف هذين البندين تم الحصول على مؤشر DTF غير ذلك دلالة، $X^2(200)$ = 208.23.

هذا ولم ينتج عن مقارنة الولايات المتحدة الأمريكية ببولندة أدائية اختبارية تفاضلية ذات دلالة، وكانت قيمة $X^2(245)$ = 272.86.

التوتر بشكل عام:

بينت المقارنات بين الولايات المتحدة والمكسيك أدائية اختبارية تفاضلية ذات دلالة: DTF = 0.012، $X^2(252)$ = 560.61، P<0.001 وتم اقتراح إزالة بند واحد هو (كثير من الأشياء شديدة التوتر) ونتج عن إزالة هذا البند مؤشر (DTF) قيمته 0.006، $X^2(252)$ = 266.15 .

وبينت المقارنات بين الولايات المتحدة والهند عدم وجود أدائية اختبارية تفاضلية ذات دلالة وكانت قيمة $X^2(200)$ = 225.63. وأخيراً نتجت عن المقارنات بين الولايات المتحدة وبولندة أدائية اختبارية تفاضلية ذات دلالة: 448.23، P<0.001، DTF = 0.074، $X^2(245)$ = وتم اقتراح استبعاد بندين وهما «مقلق» و«تحت الضغط» وعند استبعادهما من التحليل وجد مؤشر DTF غير ذي دلالة وأن قيمة $X^2(245)$ = 253.64.

التوافق بين DIF للمجموعات المتعددة وتحليلات : DTF

قبل مناقشة هذه النتائج وما تتضمنه من أجل تكييف المقاييس، من الممتع دراسة التناسق (أي التوافق) العالي النسبة بين نوعي التحليل الأدائي التفاضلي. وبشكل خاص بينت التحليلات DTF وDIF متعددة المجموعات بدقة الدرجة ذاتها في بند الرضى بالإشراف (يتدخل في عملي) وانحياز بنود الرضى بالعمل ذاته (يمثل تحدياً) و(ممل) التي ظهرت فيها منحازة بالإضافة لذلك بين بند الرضى بالعمل ذاته (هدر الوقت) DIF في كلا تحليلي DTF وDIF للمجموعات المتعددة.

وكان الفرق الوحيد بين نوعي التحليل أن ثلاثة بنود إضافية اقترحت إزالتها عند المقارنة بين الولايات المتحدة والعينة المكسيكية. أخيراً كانت نتائج مقياس التوتر بشكل عام متناسقة بين طريقتي DIF وتم التعرف على بندي (مقلقل) و(مريح) على أنهما بندان يتصفان بـ DIF في تحليل DIF متعدد المجموعات. وقد وجد أن بند (مقلقل) كان يسهم في حصول DTF بين الولايات المتحدة وبولندة وكان بند (مريح) يسهم في حصول DTF بين الولايات المتحدة والهند. إلا أن تحليل DTF أيضاً أوحى بوجوب إزالة بندين آخرين عند مقارنة الولايات المتحدة بالمكسيك والولايات المتحدة ببولندة.

المناقشة:

قمنا في هذا الفصل بوصف توجه لتاختبار التعادل بين المقاييس المكيفة وأشكالها الأصلية إجرائياً. تجدر الملاحظة أن البدائل متوفرة لكل خطوة من هذا الإجراء التحليلي. فهناك طرق متعددة لدراسة البُعدية وتقييم نواظم البنود IRT واختبار مناسبة للنموذج المقيم ووصل القياسات ومعرفة كمية حجم DIF وقد اختيرت الأساليب في تحليلنا لتبيّن كفاءتها (جيدة العلم) في الدراسات المتماثلة (مثلاً التقييم الهامشي - الوصل التكراري) أو لأنها تختبر لنا فرضية معينة تهمنا (مثل كيم وصحبه 1995) DIF متعدد المجموعات: "راجيو وصحبه (1995) وDTF" ولكن تجدر الملاحظة أن الطرق الأخرى يمكن أن تعمل أيضاً أو أنها تَفْضُلُ هذه الطرق. كما أنه لم يسبق أن أجري بحث مُمَاثل يقارن بين الفعالية النسبية لمجموعات التحليلات المطلوبة من أجل اختبار تكيفات المقاييس والاختبارات. يمكن لدراسة تماثلية واسعة أن تتقاطع عاملياً مع طرق أخرى عديدة من أجل تقييم نواظم البنود وكذلك طرق عديدة من أجل اختبار مناسبة النموذج المقيم وهناك بدائل للوصل وطرقاً لتقييم DIF. يقدم التحليل العاملي طريقة بديلة لدراسة التكيفات. وعادياً تفترض نماذج التحليل العاملي أن المتغيرات الظاهرة ترتبط خطياً

بالبنى الكامنة وأن المتغيرات الظاهرة تتبع توزيعاً متعدد الاختلافات طبيعياً (اعتيادياً). ولكن هذه الافتراضات تنتهك بشكل مزرٍ بفعل البنود ذوات الدرجات الثنائية وعلى ذلك لا يتوقع الحصول على نتائج ذات معنى من التحليل العاملي. إلا أن بنداً ذا سبع فئات في الإجابة أو حتى خمس منها يمكن أن يقدم مثالاً تقريباً جيداً على التوزيع الطبيعي في بعض الظروف (أي عندما تكرس الفئات الوسطى ولا يكرس الفئات المتطرفة إلا القليل من الناس نسبياً -انظر دراسكو دورانز 1982) في مثل هذه الحالات يمكن لتحليل سوداريوم (1974)(MACS) الذي يتناول تحليل المعدل والبنية مشتركة التباينات أن يقدم بديلاً جيداً عن طرق IRT.

كلتا الطريقتين (IRT) المستخدمة هنا و(MACS) تقدمان للباحثين أداة نافعة لأنهما تقرران دقة تمثيل الأشكال المكيفة (في المقاييس) بالتعرف على البنود التي تقيس بشكل متعادل بين الثقافات المختلفة والثقافات التي لا تعطي مثل هذه القياسات. فبفرض أنه لا توجد بنود كثيرة تبدي (DIF) يمكننا مقارنة مفهوم الرضى بالعمل ذاته مثلاً بين الثقافات باستخدام البنود التي لا تبدي (DIF) فقط. ولكن بما أن الثقافات المقارنة معاً تتزايد باضطراد، يبدو أنه من المحتمل أن يصبح عدد البنود التي لا تبدي DIF قليلاً بحيث لا يمكن إجراء مقارنة ذات معنى باستخدامها. في مثل هذه الحالة كيف يتسنى للباحثين إجراء مقارنات ذات قيمة فعلية عن الرضى بالعمل بين الثقافات؟ نقترح استخدام إجراءات تعادل الاختبارات (انظر كولين وبرينان 1995) لوصل ما نلاحظه من قياسات الدرجات إحصائياً. في هذا التحليل يمكن للباحث استخدام تحليل DIF الذي قدمه راجيو وصحبه (1995) للتعرف على المجموعة الثانوية من البنود التي تقدم قياساً معادلاً من أجل كل ثقافة مدروسة مع الثقافة المرجع. يمكن أن تعتبر هذه البند (بنوداً مشتركة) في معادلة البنود. ويمكن للمجيبين في الثقافتين أن يُعتبروا (مجموعات غير متعادلة "متساوية") والمعلومات يمكن النظر إليها على أنها تم الحصول عليها من تصميم أخذ عينات ذي البند المشترك في مجموعة غير (متعادلة متساوية) (انظر ص 18-

21 كولين وبينان 1995)، ومن هنا فأي من إجراءات التعادل من أجل هذا التصميم يمكن أن يستخدم لإعادة تدريج الدرجات من أجل النسخة المكيفة للمقياس إلى قياس لغة المصدر. بعد مثل تلك المعادلة سيكون من الممكن مقارنة مثلاً درجات الرضى بالعمل، والاستنتاج بأن العاملين في ثقافة ما أكثر رضى من العاملين في ثقافة أخرى. وبالطبع سيكون المدى الذي تمثل فيه العينات المنتقاة من ثقافات المرجع والثقافات موضوع البحث التي تمثل المجتمعات التي تؤخذ منها موضوعاً جوهرياً، ولكن معادلة الاختبارات يمكن أن تقدم طريقة لحساب درجات يكون من الممكن مقارنتها بشكل فوري.

*****

المراجع

Bach, C. L. (1998, July). U.S. international transactions. *Survey of Current Business*, pp. 47–57.

Baker, F. B. (1992). *Item response theory: Parameter estimation techniques.* New York: Marcel Dekker.

Baker, F. B., Al-Karni, A., & Al-Dosary, I. M. (1991). EQUATE: A computer program for the test characteristic curve method of IRT equating. *Applied Psychological Measurement, 15,* 78.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37,* 29–51.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46,* 443–459.

Bock, R. D. & Lieberman, M. (1970). Fitting a response model for *n* dichotomously scored items. *Psychometrika, 35,* 179–197.

Candell, G. L., & Drasgow, F. (1988). An iterative procedure for linking metrics and assessing item bias in item response theory. *Applied Psychological Measurement, 12,* 253–260.

Drasgow, F. (1989). An evaluation of marginal maximum likelihood estimation for the two-parameter logistic model. *Applied Psychological Measurement, 13,* 77–90.

Drasgow, F., & Dorans, N. J. (1982). Robustness of estimators of the squared multiple correlation and squared cross-validity coefficient to violations of multivariate normality. *Applied Psychological Measurement, 6,* 185–200.

Drasgow, F., & Hulin, C. L. (1988). Cross-cultural measurement. *Interamerican Journal of Psychology, 21,* 1–24.

Drasgow, F., Levine, M. V., Tsien, S., Williams, B., & Mead, A. D. (1995). Fitting polytomous item response theory models to multiple-choice tests. *Applied Psychological Measurement, 19,* 143–165.

Drasgow, F., & Parsons, C. K. (1983). Application of unidimensional item response theory models to multidimensional data. *Applied Psychological Measurement, 7,* 189–199.

Fleer, P. F. (1993). A Monte Carlo assessment of a new measure of item and test bias. (Doctoral dissertation, Illinois Institute of Technology, 1993). *Dissertation Abstracts International, 54-04,* 2266B.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications.* Boston: Kluwer-Nijhoff.

Hanisch, K.A. (1992). The Job Descriptive Index revisited: Questions about the question mark. *Journal of Applied Psychology, 77,* 377–382.

Hattie, J. A. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement, 9,* 139–164.

Hofstede, G. (1980). *Culture's consequences.* Beverly Hills: Sage.

Hui, C. H., & Triandis, H. C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology, 20,* 296–309.

Hulin, C. L., Drasgow, F., & Komocar, J. (1982). Application of item response theory to analysis of attitude scale translation. *Journal of Applied Psychology, 67,* 818–825.

Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). *Item response theory: Application to psychological measurement.* Homewood: Dow Jones–Irwin.

Junker, B., & Stout, W. F. (1994). Robustness of ability estimation when multiple traits are present with one trait dominant. In D. Laveault, B. Zumbo, M. Gessaroli, & M. Boss (Eds.), *Modern theories of measurement: Problems and issues* (pp. 31–61). Ottawa, Canada: Edumetrics Research Group, University of Ottawa.

Kim, S.-H., Cohen, A. S., & Park, T.-H. (1995). Detection of differential item functioning in multiple groups. *Journal of Educational Measurement, 32,* 261–276.

Kolen, M. J., & Brennan, R. L. (1995). *Test equating: Methods and practices.* New York: Springer.

Landar, H. J., Ervin, S. M., & Horowitz, A. E. (1960). Navaho color categories. *Language, 36,* 368–382.

Leung, K., & Drasgow, F. (1986). Relation between self-esteem and delinquent behavior in three ethnic groups: An application of item response theory. *Journal of Cross-Cultural Psychology, 17,* 151–167.

Lord, F. M. (1977). A study of item bias, using item characteristic curve theory. In Y. H. Poortinga (Ed.), *Basic problems in cross-cultural psychology* (pp. 19–29). Amsterdam, Netherlands: Swets & Zeitlinger.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Marin, G., Gamba, R. J., & Marin, G. V. (1992). Extreme response style and acquiescence among Hispanics. *Journal of Cross-Cultural Psychology, 23,* 498–509.

McLaughlin, M. E., & Drasgow, F. (1987). Lord's chi-squared test of item bias with estimated and known person parameters. *Applied Psychological Measurement, 11,* 161–173.

Mislevy, R. J., & Bock, R. D. (1989). *PC-BILOG 3.* Mooresville: Scientific Software.

Raju, N. S., van der Linden, W. J., & Fleer, P. F. (1995). IRT-based internal measures of differential functioning of items and tests. *Applied Psychological Measurement, 19,* 353–368.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics, 4,* 207–230.

Richards, I. (1953). Toward a theory of translation: Studies in Chinese thought. *American Anthropological Association, 55* (Memoir 75). Chicago: University of Chicago Press.

Roznowski, M. (1989). An examination of the measurement properties of the Job Descriptive Index with experimental items. *Journal of Applied Psychology, 74,* 805–814.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement,* No. 17.

Segall, D. O. (1983). *Assessment and comparison of techniques for transforming parameters to a common metric in item response theory.* Doctoral dissertation, University of Illinois at Urbana-Champaign.

Smith, P. C., Kendall, L., & Hulin, C. L. (1969). *The measurement of satisfaction in work and retirement.* Chicago: Rand McNally.

Smith, P. C., Sademan, B., & McCrary, L. (1992, May). *Development and validation of the Stress in General (SIG) scale.* Paper presented at the meeting of the Society for Industrial and Organizational Psychology, Montreal, Canada.

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *Psychometrika, 27,* 229–239.

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7,* 201–210.

Stone, C. A. (2003). Empirical power and type I error rates for an IRT fit statistic that considers the precision of ability estimates. *Educational and Psychological Measurement, 63*(4), 566–583.

Stone, C. A., & Hansen, M. A. (2000). The effect of errors in estimating ability on goodness-of-fit tests for IRT models. *Educational and Psychological Measurement, 60*(6), 974–991.

Triandis, H. C. (1972). *The analysis of subjective culture.* New York: Wiley.

Triandis, H. C. (1990). Cross-cultural studies of individualism and collectivism. In J. Berman (Ed.), *Nebraska Symposium on Motivation, 1989* (pp. 41–133). Lincoln: University of Nebraska Press.

Triandis, H. C. (1994). *Culture and social behavior.* New York: McGraw-Hill.

Triandis, H. C. (1995). *Individualism and collectivism.* Boulder: Westview Press.